A Container-based Monitoring Approach for Auditability of Cloud Providers Using Blockchain

***Abstract*—With the rapid development of cloud-based systems providing numerous pay-as-you-go and elastic services, we are increasingly exposed to data leakages and incidents of unauthorised access to customers' personal data. In this context, cloud customers should be able to choose available services based on their security and privacy requirements. This paper presents the design of a container-based monitoring architecture for multi-cloud ecosystems which produces an audit trail of providers' operations on customers data in a secure and automatic way. The architecture makes use of Blockchain technology to automatically verify compliance with the data protection, data minimisation and data transfer obligations contemplated in the General Data Protection Regulation (GDPR) by cloud providers' activities that are recorded using the monitoring system. We propose three smart contracts in order to: (i) enable customers to give a preference for the verification of such GDPR obligations in their selected services, (ii) store the information captured via the container using the monitoring system, and (iii) verify compliance with the GDPR by cloud providers. The smart contracts are deployed in a Blockchain test network to determine their costs.**

***Index Terms*—Blockchain, Container-based monitoring, Data privacy, General data protection regulation, Smart contracts.**

I. Introduction

Driven by lower costs of ownership, elastic on-demand services, enhanced interoperability and the insights derived from machine learning, cloud computing synthesises the best of previous mainframe and personal computer paradigms. Given the manifold benefits of cloud computing, companies increasingly rely on cloud vendors' servers and infrastructure to host and operate their websites and mobile apps. Cloud platform services are also gradually becoming the preferred choice for developers to create and deploy middleware and other customised applications. Similarly, cloud-based applications, which allow users to run software through web browsers without the need to install any specific software, are growing in popularity. As a result, data is migrating to the cloud, a trend that is unlikely to be halted or reversed in the foreseeable future [2].

Aside from its undeniable benefits, however, cloud computing has a negative flipside. The growing amounts of data stored in the cloud, coupled with the complexity of cloud ecosystems, raise significant governance and compliance concerns under the recently introduced EU General Data Protection Regulation (GDPR) [3]. GDPR compliance concerns serve as a barrier to migration to the cloud, especially for SMEs, which are more likely to "find it difficult to exercise the full control required by data protection legislation on how the provider delivers the requested services" [1].

Cloud-based solutions are typically "layered", involving a complex and convoluted chain of cloud service providers. In this architecture, it can be difficult for the data controller to effectively check the cloud vendors' data handling practices and thus to be certain that the data is being handled in a lawful manner. In the same vein, due to loss of governance and control, the cloud customer may be incapable of providing evidence of compliance with *inter alia* the data location and transfer, "privacy-by-design" and accountability requirements laid down in the GDPR (Arts. 44–47, 25 and 5(2)) [8].

On the other hand, end-users (i.e. "data subjects" in the GDPR parlance) are seldom aware of the highly intricate and layered architecture of cloud computing [4]. They typically interact only with a Web interface rather than the larger, composite ecosystem, entrusting their personal data and identity to the consumer-facing component without realising that the cloud-based application may share their data with several back-end services (e.g. providers of cloud-hosted analytics and online advertising). In this opaque context, it is hard for data subjects to exert any control over their personal data after the initial disclosure. Also, when multiple subcontractors are involved in a cloud solution, the risk of personal data being processed for further, incompatible purposes is quite high. [1] Crucially, the dearth of transparency that characterises cloud ecosystems impede the data subject's ability to give "informed consent" to the use, sharing, and repurposing of their personal data. Users experience a lack of control over their personal data, exacerbated by the growing number of data breaches resulting from cyberattacks on cloud vendors. This threatens to undermine consumer trust and hinder the uptake of cloud computing.

To realise the full potential and benefits of cloud computing, the data protection concerns above must be effectively addressed. To this end, cloud ecosystems should be both transparent and auditable. In particular, data subjects should be able to verify who processes their personal data, with whom it is shared, what are the legal bases justifying the processing, and crucially, whether their consent given for one or more specific purposes is respected or overridden. In turn, cloud customers, acting as data controllers, should be able to check whether their cloud service providers are processing their users' personal data in a lawful manner, in accordance with their instructions, so they can both be sure that their users' data protection rights are being respected and be able to prove compliance with the GDPR obligations and requirements to which they are bound.

Blockchain technology is becoming a promising solution

for enhancing data privacy in cloud environments [5], [7]. It enables the production of an audit trail of cloud providers through a fully distributed, provable secure and consensus-based way [9]. For instance, an automatic way for tracking and enforcing data sharing agreements between a customer and cloud providers with the aid of smart contract and Blockchain was proposed in [10]. In this approach, the violation of the shared agreements by the providers were detected by a number of voters listed in a voting contract. Similarly, the integration of Blockchain and GDPR resulted into the design of a privacy-aware architecture for cloud ecosystems - promoting access control - in [11]. Relatedly, in [12], a number of GDPR rules were translated into smart contracts in order to automatically verify legal compliance in the operations executed by providers on cloud customers' data. Although the foregoing approaches make use of the GDPR and Blockchain for improving data privacy, none of them considers the preferences of cloud customers for verifying GDPR obligations. Moreover, they did not technically examine how container technology is used within a privacy-aware cloud architecture in order to monitor the activities of cloud providers and improve customers trust. In order to address this, we propose a new container-based architecture making use of Blockchain and GDPR. The key contributions of this parer are summarized below:

- an online pharmacy scenario is presented to show GDPR concerns within cloud composite services;
- a container-based monitoring framework is proposed to record the operations executed on personal data, which is exchanged or stored while handling the customer query for accessing the online pharmacy application hosted in a multi-cloud environment;
- three smart contracts are proposed to verify the operations of cloud providers on personal data in the light of three GDPR obligations (data protection, data minimisation and data transfer);
- our Blockchain-based solution allow cloud customers to give their priorities for verifying compliance with the three GDPR obligations above;
- the proposed smart contracts are deployed in a Blockchain test network for evaluating their costs and mining time.

The remainder of the paper is structured as follows: Section II presents the online pharmacy scenario. Section III presents the design of the container-based architecture and proposes the smart contracts. Section IV describes experimental results of our Blockchain-based technique, and finally Section V sets forth a number of conclusions.

II. An Online Pharmacy

Imagine a patient who wishes to buy drugs available on prescription online. The patient visits an online pharmacy to place the order, make the payment and get his prescription shipped to his home address. Online transactions like this seem very simple in the eye of consumers, but there is much going on behind the curtains. The online pharmacy is actually a multi-layered cloud-based service involving multiple flows of personal data from the user to different service providers.

Upon the patient placing the order, the pharmacy accesses the patient's personal data, including name, address details, age, general practitioner (GP) diagnosis, electronic version of the prescription, and bank account details. Moreover, the pharmacy maintains an electronic health record (EHR) to store useful information about the medical status of patients. The online pharmacy uses a Russia-based IaaS vendor (Cloud4U) to host and operate its website and mobile app. All the data above are transferred to Cloud4U's servers located in Moscow.

The pharmacy's website and mobile app are embedded with so-called "social plugins" (a "Like" button and a "Share" button) from a leading social network (Friendface). The default implementation of Friendface's APIs and SKDs is designed to automatically transmit data from the online pharmacy to Friendface the moment a user visits the pharmacy's website or opens the app. The automatically transmitted data includes events data (such as "App installed" and "SDK initialised") and location data. Friendface uses this data to enrich the patient's profile it has built over time (i.e. profiling), which is valuable for its ad targeting business.

The online pharmacy uses the real-time bidding (RTB) system of a prominent online advertiser and intermediary (Froogle) to sell advertising inventory space on its website and mobile app, and thus derive another revenue stream. Every time a user visits the pharmacy's website or uses the pharmacy's mobile app, RTB cookies and tracking pixels are placed on the user's device, thereby enabling the broadcast of said user's personal data to hundreds of companies in the ad tech chain. The personal data includes the user's location, device description, unique tracking ID, IP address, data broker ID segment, and what the user is currently reading or watching, among other data. The broadcast of personal data is made by Froogle's "Supply Side Platform" on behalf of the pharmacy, with an aim to solicit bids from companies which may want to show an ad to the pharmacy's user.

Lastly, the pharmacy subcontracts payment and shipping service providers to handle the payment and delivery of medicines. The payment service provider receives the patient's name and bank account details from the pharmacy, offering two alternatives, Western Union and Paypal, to manage the payment process. The shipping service provider, in turn, receives the patient's name and address details from the pharmacy, whereupon it packs the order, sends the patient a reference number to track his parcel, and delivers it.

Each data flow in this cloud-based ecosystem amounts to a "personal data processing" operation, thus triggering the application of the GDPR. Within this regulatory framework, the online pharmacy is the data controller, as it determines the purposes and means of the personal data processing. Specifically, it processes the patient's personal data to process and complete the order (purpose), and it does so through the multi-layered architecture explained above (means). Cloud4U and the payment and shipping providers, conversely, are data processors, as they process the patient's personal data on

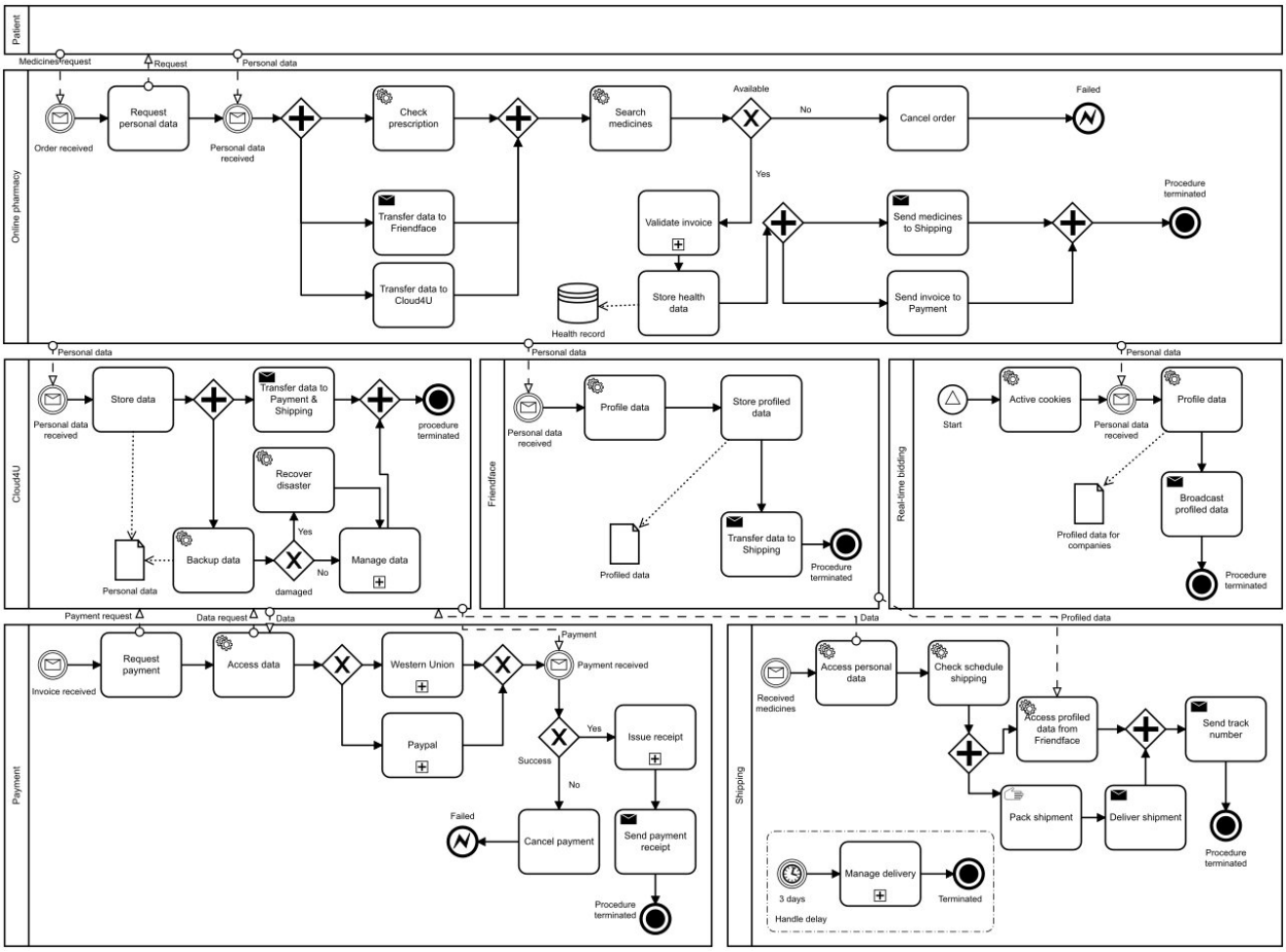

Fig. 1. A business process model for online pharmacy

behalf of the controller, in furtherance of the purpose the controller has determined. Friendface and Froogle, in turn, are joint controllers with the online pharmacy in respect of the operations involving the collection and disclosure of the patient's personal data to the first two companies, and sole controllers in respect of the operations involving data processing carried out by them after the patient's personal data have been transferred. Determining these roles is an important preliminary step, as controllers and processors are subject to different obligations. For example, data protection rights can be enforced against controllers only.

Although the GDPR subsumes all operations performed on personal data within the same term (i.e. processing), for the purposes of our architecture we identify and name three separate operations. Firstly, order processing, which encompasses the provision of personal data by the patient to the pharmacy upon placing the order, and the disclosure of personal data by the pharmacy to the payment and shipping providers to process payment and delivery. Secondly, international transfer, which refers to the transfer of personal data from the online pharmacy to actors located outside the EU. Thirdly, online advertising, which include the collection and processing of personal data by Friendface and Froogle for the provision of their online advertising services.

There are many data protection issues concerning the three operations above. For the sake of simplicity and on account of space limitations, however, we will focus on the following three:

- Transfers of personal data to a non-EU country or international organisation are restricted. There are two main ways of allowing international transfers of personal data: on the basis of an adequacy decision by the European Commission, or in lieu thereof, where the controller or processor provides appropriate safeguards, including enforceable rights and legal remedies for the data subject [8], Art. 45. These appropriate safeguards can be established by a legally binding and enforceable instrument between public authorities or bodies; binding corporate rules (BCRs); standard data protection clauses adopted either by the European Commission or by a supervisory authority; codes of conduct; or certification mechanisms [8], Arts. 46 and 47.
- The principle of data security requires that appropriate technical or organisational measures are implemented

when processing personal data to protect the data against accidental, unauthorised or unlawful access, use, modification, disclosure, loss, destruction or damage [8], Arts. 5(1)(f) and 32(1). Depending on the specific circumstances of each case, these measures may include, for example, pseudonymising and encrypting personal data.[1]

- Under the data minimisation principle, the personal data being processed must be limited to what is necessary [8], Art. 5(1)(c). Whether the data is necessary will depend on the controller's specified purpose for collecting and using the personal data.

III. Architecture

We propose a multi-tier container-based monitoring architecture as shown in Fig. 2, based on the above online pharmacy case study. In this architecture, the operations related to the online pharmacy application are executed inside a container hosted over multiple cloud service providers (at Infrastructure as a Service layer).

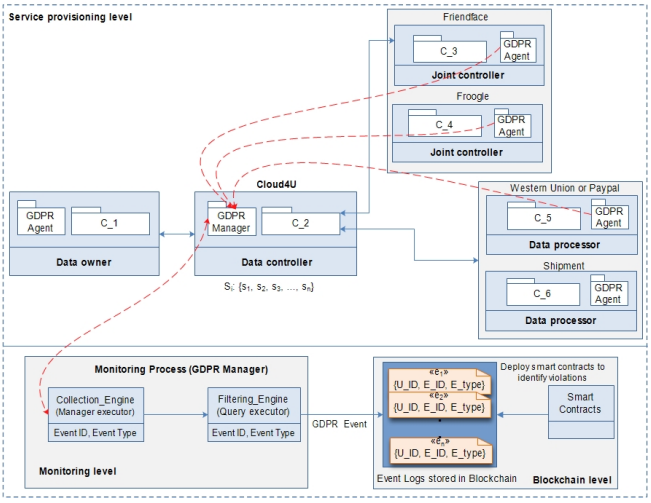

Fig. 2. Multi-tier Cloud Monitoring Architecture

The proposed architecture consists of three functional levels: 1) service provisioning, 2) monitoring, and 3) blockchain level. At the service provisioning level, whenever any customer (i.e. a patient) places an order on the online pharmacy website/app hosted over the cloud, the data exchange between the customer and the cloud provider happens in a containerised environment. The controller[2] cloud provider (Cloud4U) receives the request and distributes the service compositions to different processor cloud providers (Western Union/Paypal or shipment provider). Alongside, there may be a possibility of some automated joint controllers (FriendFace and Froogle) coming into actions due to default SDKs or plugins embedded within the controller cloud provider's perimeter. Once the service is initiated, the monitoring system is activated. The monitoring system contains two main components, 1) GDPR-Agent, and 2) GDPR-Manager, which is responsible for monitoring the data operations taking place on the data in question throughout its lifetime and recording them on the blockchain. At the blockchain level, the smart contracts are deployed to analyse potential GDPR violations.

A. Monitoring System

The GDPR-Agent is a monitoring agent which is executed inside a container to track the data operations in the container-hosted online services. The GDPR-Agent understands the underlying heterogeneity of the containers deployed on a multi-tier cloud infrastructure. The main task of the GDPR-Agent involves the collection of data operation statistics along with multi-level statistics and its transmission to the GDPR-Manager. The GDPR-Manager runs inside the controller cloud provider, and collects all the data operations monitored by different GDPR-Agents deployed in the entire service chain. The working of these components is described below.

1) GDPR-Agent: GDPR-Agent is a software element that when activated collects the events related to the data operations being performed inside a container in a cloud-hosted service (like the online pharmacy). The GDPR-Agent also known as *SmartAgent* receive data from various sub agents (like *Process Agent*, *System Agent*, *File Agent*, *User Access Agent*, and *Network Agent*), which are shown in Fig. 3.

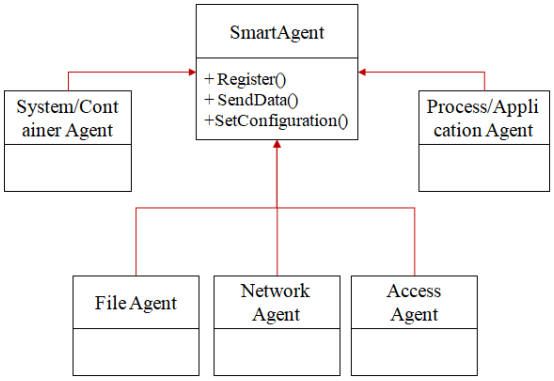

Fig. 3. GDPR-Agent model

There are three operations related to the GDPR-Agent: 1) a registration request is sent by the agent to the manager, 2) data related to operations and other statistics are sent by the agent to the manager periodically, and 3) the agent sends its configuration information to the manager, which can update the configuration parameters of the agent if required. Initially, this agent has to register with the GDPR-Manager according to the process shown in Fig. 4

2) GDPR-Manager: The GDPR-Manager receives the information related to the data operations monitored by different GDPR-Agents (deployed inside containers). The entire com-

[1]Given that 'data security' has in computer science a meaning other than that contemplated in the GDPR, in this paper we refer to this obligation as 'data protection', a term which we believe encapsulates better the essence of this obligation

[2]Legally speaking, Cloud4U is a processor, but for the purposes of the architecture, the infrastructure that hosts the controller (i.e. the online pharmacy) is the controller

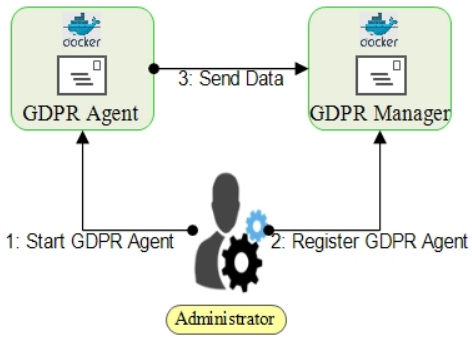

Fig. 4. Registration and data collection of GDPR Agent

munication process between GDPR-Agent and the manager is shown in Fig. 5.

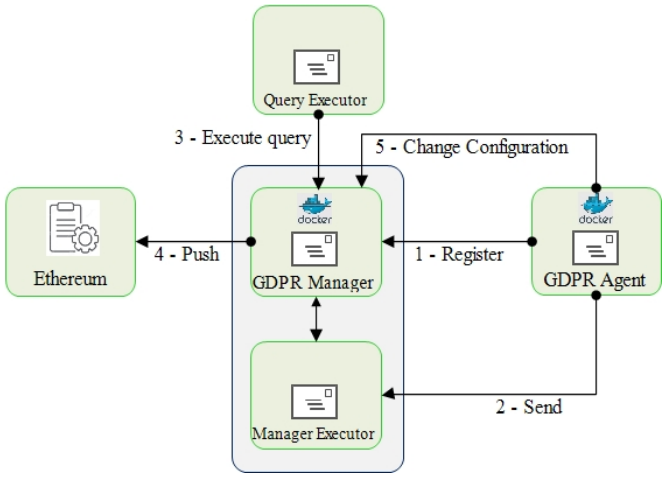

Fig. 5. Communication process

Once the GDPR-Agent is registered with the GDPR-Manager (step 1), the access key and the endpoints are sent to the GDPR-Agent by the manager. Thereafter, the Agent sends the monitored events to the GDPR-Manager through the Manager Executor (step 2). The GDRP-Manager communicates with the GDPR-Agent using Advanced Message Queuing Protocol (AMQP). The data operations and statistics (such as system, network, file, process and user access) collected from the container are added to multiple queues via AMQP messages, where they are filtered to extract GDPR-relevant metrics. The data received from the monitoring agent is exposed to rule-based queries activated inside a filtering engine through a query executor. These queries filter out the GDPR-specific metrics from the overall statistics and data collected by the GDPR-Agents (step 3). After this, the GDPR-Manager records the GDPR metrics onto the Ethereum blockchain using Push-based mechanism (step 4). Finally, the configuration information is sent to the manager, which can update the same if required (dynamic configuration enables real-time communication) (step 5). To submit the GDPR metrics to the Ethereum blockchain, the GDPR-Manager has to create a ethereum account and get an ETHER ID. This process is shown in Fig. 6.

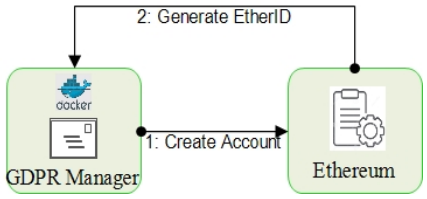

Fig. 6. Registration process of GDPR Manager

3) Monitoring Process: The monitoring process involves different steps and programming operations as shown in Fig. 7. These steps and related operations are described below.

- All the GDPR-Agents are deployed in the container of the user and cloud providers; the data operations are recorded and sent to the GDPR-Manager.
- These events or metrics are added into different queues based on their type or characteristic via AMQP messages using a Publish/Subscribe mechanism. For this purpose, AMQP producers compatible with RabbitMQ were built, which carry out the publication of the data operations recorded by the GDPR-Agent. Now, RabbitMQ was the AMQP server selected to act as a message-broker between the GDPR-Agent and the GDPR-Manager. For example, if an agent monitors the transfer operation, then its AMQP client triggers a connection to the RabbitMQ server and send the recorded data to the "Process" queue.
- The data sent to the GDPR-Manager is collected through a manager executor working in the collection engine.
- The data collected at the collection engine is further passed on to the filtration engine where the query executor triggers a rule-based query to filter the GDPR metrics. This is done to avoid the addition of non-GDPR event in the blockchain. This enhances the efficiency of the verification process.
- Finally, this data is filtered and added to the blockchain as a transaction. For this reason, the manager (which already has an Ethereum account) creates a genesis (i.e. first) transaction to which data, the timestamp, the signature, and the public key of the controller cloud provider are appended, as shown below.

$$T1 = [Data, TimeStamp, sign, Publickey]$$

- This is followed by the creation of the genesis block (B) with a unique ID; B: $[B_{ID}]$
- The genesis transaction (T1) is added to the block (B) after validation by the trusted nodes.
- The user creates new transactions in the same manner until the block achieves the maximum block size (fixed) or threshold (variable).

$$B : [T1, T2, T3, \ldots, Tn]$$

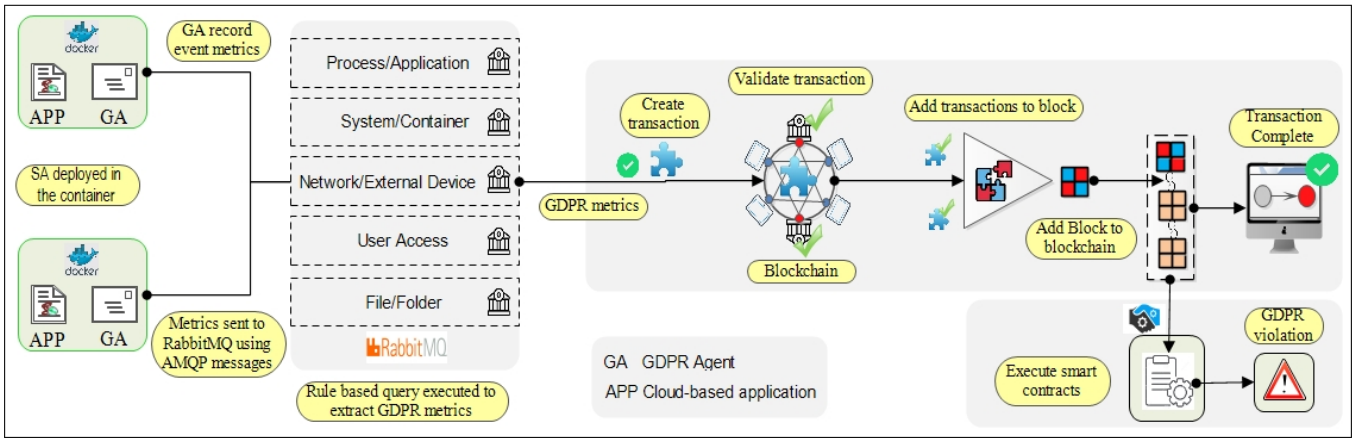

Fig. 7. Architecture

- Finally, the block is added to the blockchain where the smart contracts are deployed to verify the data operations according to GDPR.

B. GDPR-supported smart contracts

Figure 8 represents our proposed smart contracts. Two smart contracts, namely *GDPR-priority* and *container-log*, are proposed to record the information required for checking GDPR compliance. *verification* is deployed to verify compliance with the aforementioned obligations (data protection, data minimisation and data transfer) by cloud providers. The activators of *GDPR-priority*, *container-log* and *verification* are respectively the data subject (user), the smart manager, and the verifier. Note that *verifier* is a trusted third party connected to the Blockchain in charge of flagging GDPR violations.

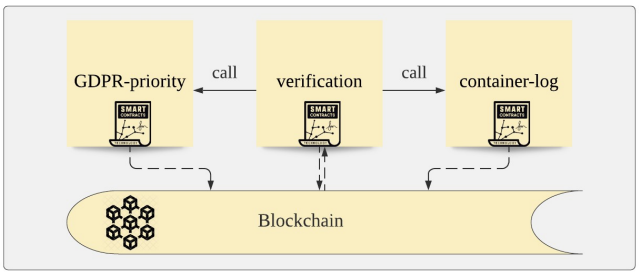

Fig. 8. GDPR-supported smart contracts

C. Determining GDPR Compliance Preference

Verifying compliance with GDPR obligations (i.e. data protection, data minimisation and data transfer) is a costly process. In some cases, a data subject may not be concerned about the observance of the obligations above, and thus prefer not to verify compliance therewith and save money. This phase enables data subjects to activate the *GDPR-priority* contract in order to specify their preferences for verifying compliance with obligations. The smart contract allows the data subject to give a compliance score, thus signalling his preference in respect of each obligation.

Definition 1. Let Σ be a set of GDPR obligations. A compliance score is a function $\mathcal{S} : \Sigma \mapsto [0, 1]$. A compliance score $\mathcal{S}(\sigma) = 1$ shows a must for the verification of an obligation $\sigma \in \Sigma$ and $\mathcal{S}(\sigma) = 0$ ignores the verification of σ.

The contract stores such scores into a Blockchain to inform the verifier about the data subject's preferences concerning the verification of GDPR obligations. For example, *data protection* is likely to be the greatest concern of the customers of the online pharmacy depicted in Fig. 1, since their personal data involve healthcare information, which is categorised as sensitive.

D. Recording Data Processing Operations

This phase collects and sends a number of useful information for verifying GDPR compliance by cloud providers. The GDPR-manager of the container deploys the *container-log* smart contract to send such information to a Blockchain. The information includes (i) the provider's address (p), (ii) processed operations (A_p) on user data, (iii) processed personal data items (D_p), (iv) collected personal data items from user (D_{c_p}), (v) encryption status of operations ($\mathcal{E}_{a_p}$), (vi) the country name of the provider (loc_p).

Thereafter, the data operations are recorded by the GDPR-Agent; the filtering engine is triggered to extract only the events that meet the GDPR obligations defined by Σ.

Definition 2. Let $\mathcal{E}$ be a set of data operations and multi-level statistics recorded by the GDPR-Agent. The data operations ($\mathcal{E}_\Sigma$) based on GDPR obligations Σ are filtered using the filtering function $\mathcal{F} : \mathcal{E} \mapsto [0, 1]$, so that

$$\mathcal{E}_\Sigma = \begin{cases} 1: & \text{If } \mathcal{E} \in \{\text{read, write, transfer}\} \\ 0: & \text{Otherwise} \end{cases}$$

E. Verifying GDPR Compliance

This phase verifies the operations of cloud providers (data controllers/processors) executed on personal data in the light of the GDPR obligations chosen or prioritised by the data subject. The assumption is that the verifier determines a threshold θ_σ

for verifying compliance with a specific obligation σ. Such a threshold is subjective and denotes the interest of the verifier in checking GDPR compliance.

Definition 3. A preference for checking obligations is a strict partial order, denoted by $\mathcal{P} = (\Sigma, <^{\mathcal{P}})$, where $<^{\mathcal{P}} \subseteq \Sigma \times \Sigma$. If $\sigma, \sigma' \in \Sigma$ are two different GDPR obligations, then $\sigma <^{\mathcal{P}} \sigma'$ is expressed as "σ' is preferred rather than σ".

Given a set of obligations selected by the data subject and their verification time, the following definition gives the verifier a priority for detecting violations in accordance with the preferences of the data subject.

Definition 4. Let $\sigma, \sigma' \in \Sigma$ be two GDPR obligations, θ_σ and $\theta_{\sigma'}$ be the thresholds predefined by the verifier, $\mathcal{S}(\sigma)$ and $\mathcal{S}(\sigma')$ be the compliance scores determined by the data subject, so that $\mathcal{S}(\sigma) \geq \theta_\sigma$ and $\mathcal{S}(\sigma') \geq \theta_{\sigma'}$. Checking the obligation σ' is preferred over checking σ (i.e. $\sigma <^{\mathcal{P}} \sigma'$) *iff* $\mathcal{S}(\sigma) < \mathcal{S}(\sigma')$.

Algorithm 1 Checking GDPR compliance

1: $V \leftarrow \emptyset$
2: **case** $data\ protection$
3: $\quad V \leftarrow V \cup \{p \in P \mid \exists\ a_p \in A_p \ \ and \ \ \mathcal{E}_{a_p} = \perp\}$
4: **case** $data\ minimisation$
5: $\quad V \leftarrow V \cup \{p \in P \mid D_{c_p} \not\subseteq D_p\}$
6: **case** $data\ transfer$
7: $\quad V \leftarrow V \cup \{p \in P \mid \exists\ r_p \in P \ \ and \ \ loc_{r_p} \notin BCR\}$
8: **return** V

The verifier deploys the *verification* contract to report providers breaching the aforementioned obligations (i.e. data protection, data minimisation and data transfer). The contract involves a function presented in Alg. 1 to flag GDPR violators that are extracted from a set denoted by V.

Data protection case: A provider p from the set of cloud providers P, executing a set of operations A_p on user data, is a violator if at least has an operation a_p that does not encrypt personal data.[3]

Data minimisation case: A provider p is reported as a violator, if the data items' set D_p processed by p is a subset of the data items' set D_{c_p} requested by p. In other word, a provider who collects data that is not used for processing is classified as a violator.[4]

Data transfer case: A provider p is reported as a violator, if personal data is transferred to a provider r_p such that the country of r_p (denoted by loc_{r_p}) is not included in the *BCR* set. The set includes the names of non-EU countries having an adequacy decision with the European Commission, and the names of groups of undertakings or enterprises which adhere to duly approved data protection policies to move personal data internationally over different jurisdictions [6]. Note that the encryption of the data transfer operation is also checked through the contract.

[3]For the sake of simplicity, we assume that *encryption* is the only mechanism available for protecting personal data.

[4]We assume that collected yet not processed data is not necessary for the purposes determined by the controller which justify the processing of personal data.

TABLE I
THE TRANSACTION COSTS OF SERVICE PACKAGES

	Service pack 1	Service pack 2	Service pack 3
Number of actors	2	4	6
Number of operations	9	16	23
Total container-log (wei)	1562456	2782678	3882146
Data protection (wei)	297628	743436	1401864
Data minimisation (wei)	905648	1582621	2305178
Data transfer (wei)	323501	1112821	1803427

IV. Experimental Results

A prototype has been developed via Ropsten [14], which is a public Blockchain test network. The prototype provided some details about consumed gas when executing transactions and the time the mining process takes. Our proposed smart contracts were implemented on Ethereum [13] using Solidity language. We tried to write our smart contracts with a minimum gas consumption for each function. The contracts were compiled by Remix, a browser-based compiler for Solidity codes written in Ethereum virtual machine. The smart contracts *GDPR-priority*, *container-log* and *verification* were executed in the Ropsten network. The amount of gas consumed for deploying a contract was 315846 for *GDPR-priority*, 1146109 for *container-log* and 1453649 for *Verification*.

A. Transaction costs

This evaluation calculates the amount of gas consumed in the execution of our proposed smart contracts. It assumes that we have three cloud service packages for the pharmacy scenario represented in Sect. II. Service package 1 includes *online pharmacy* and *payment* providers (actors). Service package 2 involves *online pharmacy*, *cloud4u*, *friendface* and *real-time binding* providers. Service package 3 includes all the six service providers depicted in Fig 1. The number of operations executed by the actors on user data is 9, 16 and 23 for service package 1, package 2 and package 3, respectively. The proposed smart contracts were deployed in the Ropsten test network and were executed five times to calculate the average results. We investigated the average cost for recording information via the container-log contract and the average costs for verifying compliance with the data protection, data minimisation and data transfer obligations. Table I shows the experimental results. As can be seen, when the number of operations/ actors increases, the amount of gas consumption increases sharply. In this experiment, the amount of gas used for recording user preference via the GDPR-priority contract was 83791 wei.

B. Verification cost vs mining time

This experiment evaluates the verification of compliance with the data protection, data minimization and data transfer obligations under different gas prices. It assumes that we have three cloud service packages, being the same as those described in Sect. IV-A. The gas price units requested by

TABLE II
THE VERIFICATION COSTS AND AVERAGE MINING TIME

	Service pack 1	Service pack 2	Service pack 3
Gas price (gwei)	45	57	70
Data protection ($)	3.12	9.70	23.19
Data minimisation ($)	9.49	20.65	38.13
Data transfer ($)	3.40	14.52	29.83
Mining time (seconds)	257	115	26

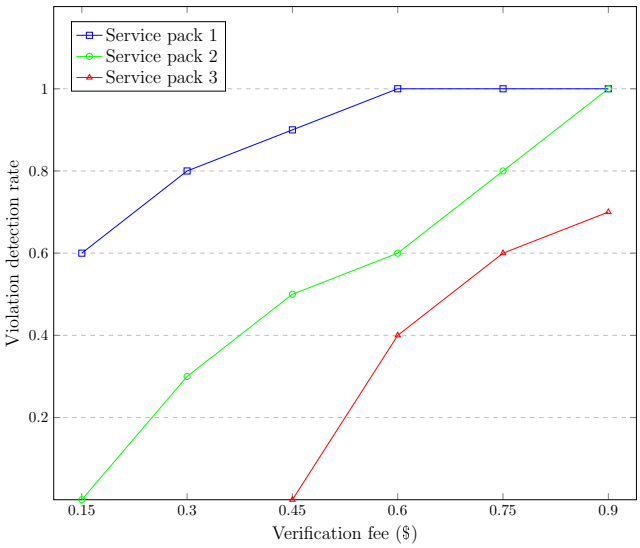

Fig. 9. Transaction fee vs Violation detection rate

user are 45, 57 and 70 *gwei* for running the transactions of service package 1, package 2 and package 3, respectively. Our proposed smart contracts were executed five times in Ropsten to calculate the average results.[5] As seen in Table II, the verification costs are expressed in USD, and the average time taken for mining such transactions is calculated in seconds. Given a service package, *data protection* has the minimum verification cost, as it only considers the encryption status of operations and hence its complexity is lesser than other obligations. In contrast, the most expensive verification cost is allocated to *data minimisation*, since not only does it deal with checking operations, but also examines the personal data processed by them. The evaluation also showed that, when the rate of gas price unit increases, there is a considerable decrease in the average mining time.

C. Evaluation of violation detection rate

This experiment evaluates the relationship between the cost paid by the patient for verifying GDPR obligations and the average rate of successful violation detection. It assumes that the aforementioned service packages are offered to the patient. We used Ropsten for executing the *verification* contract and the rate of gas price is 1 gwei. Per each execution there is a GDPR violation in a service package selected by the patient. The violation is randomly selected from amongst the data obligation, data minimization and data transfer obligations. The contract was executed ten times to calculate the average detection rate. Figure 9 illustrates the results of the experiment, where the x-axis shows the fee paid by the patient for verifying compliance with obligations and the y-axis shows the number of successfully detected violations. As can be seen, for a given price, the service package 3, involving the highest number of operations, has the lowest likelihood of violation detection. For instance, GDPR compliance cannot be detected when patients select service package 3 and their budget is $0.3.

[5]https://ethgasstation.info/

V. CONCLUSION

This paper presented the design of a container-based architecture for the purpose of improving data privacy in cloud ecosystems. The architecture makes use of Blochchain technology and supports GDPR obligations in the tracking of activities executed by cloud providers on user data. By proposing the *GDPR-priority*, *container-log* and *verification* smart contracts, the paper verified compliance with three GDPR obligations by cloud providers, namely data protection, data minimisation and data transfer. The *GDPR-priority* contract allows the user to give a preference for verifying compliance with an obligation. The *container-log* contract allows the GDPR-manager within a container to send information relevant to the aforementioned obligations to a Blockchain. The *verification* contract enables the verifier to check compliance with the obligations based on the preferences determined by the user. These smart contracts were deployed in the Ropsten test network, and the evaluation results showed that as the number of operations in a composite service grows, the verification cost increases significantly.

The architecture proposed in this paper is a promising solution to facilitate effective compliance with the GDPR, thereby protecting individuals' informational privacy and autonomy. In particular, the tamper-proof nature of the Blockchain ensures that data controllers have at their disposal a reliable audit log to verify whether their cloud vendors are processing their users' personal data in a lawful manner, in accordance with their instructions. At the same time, the audit log can prove invaluable in potential investigations of GDPR violations conducted by supervisory authorities. Moreover, the technological solution enables data subjects to verify whether the cloud-based applications they use are consistent with their privacy preferences, as well as whether the providers of those applications live up to the commitments they make in their privacy policies. Lastly, by enhancing transparency and accountability, the technological solution aligns cloud-based architectures with the data protection-by-design and by-default requirements laid down in the GDPR.

However, a work of caution is in order. There is a significant operational and conceptual gap between law and technology. The "ways of working" of each field differ significantly. Data protection law concepts such as *personal data*, *processing*, *consent*, *purpose limitation* and *legitimate interests*, to name a few, are overly broad, highly abstract, or involve an intricate substantive assessment. Computer science and software

programming, on the other hand, rely on a concise, binary, "if/then" type of language that fails to catch legal intricacies. Additional interdisciplinary research is required to broaden the set of legal rules the compliance with which can be verify through smart contracts.

Future work will focus on investigating the performance of our proposed architecture in a real cloud-based platform. The translation of more GDPR provisions into smart contracts remains another potential research avenue for future consideration.

REFERENCES

[1] Article 29 Data Protection Working Party Opinion 5/2012 on Cloud Computing 01037/12/EN WP 196, 2012.

[2] K. Bila, O. Khali, A. Erbad, and S. U. Kha, Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers, *Computer Networks*, vol. 130, pp. 94–120, 2018.

[3] B. Russo, L. Valle, G. Bonzagni, D. Locatello, M. Pancaldi, and D. Tosi, Cloud Computing and the New EU General Data Protection Regulation, *IEEE Cloud Computing*, vol. 5, no. 6, pp. 58–68, 2018.

[4] M. Virvou and E. Mougiakou, Based on GDPR privacy in UML: case of e-learning program, *in Proc. of the 8th International Conference on Information, Intelligence, Systems & Applications*, Larnaca, Cyprus, 2017.

[5] N. Al-Zaben, M. M. H. Onik, J. Yang, N-Y. Lee, and C-S. Kim, General data protection regulation complied Blockchain architecture for personally identifiable information management, *in Proc. of the International Conference on Computing, Electronics & Communications Engineering*, Southend, UK, 2018, pp. 77–82.

[6] M. Corrales, P. Jurcys and G. Kousiouris, "Smart contracts and smart disclosure: Coding a GDPR compliance framework," SSRN Electronic Journal,2018.

[7] E. Gaetani, L. Aniello, R. Baldoni, F. Lombardi, A. Margheri, and V. Sassone, Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments, *in Proc. of the First Italian Conference on Cybersecurity*, Venice, Italy, 2017, pp. 146–155.

[8] Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, 'GDPR') OJ L119/1 2016.

[9] Y. Zhang, S. Wu, B. Jin, and J. Du, A Blockchain-based Process Provenance for Cloud Forensics, *in Proc. of the 3rd IEEE International Conference on Computer and Communications*, Chengdu, China, 2017, pp. 2470–2473.

[10] H. Desai, K. Liu, M. Kantarcioglu, and L. Kagal, Enforceable Data Sharing Agreements Using Smart Contracts, *arXiv:1804.10645v1*[*cs.CY*], 2018.

[11] M. Barati and O. Rana, Privacy-aware cloud ecosystems: Architecture and performance, *Concurrency and Computation: Practice and Experience*, 2020. DOI: 10.1002/cpe.5852

[12] M. Barati and O. Rana, Tracking GDPR Compliance in Cloud-based Service Delivery, *IEEE Transactions on Services Computing*, 2020. DOI: 10.1109/TSC.2020.2999559

[13] G. Wood, Ethereum: A secure decentralised generalised transaction ledger, *Ethereum Project Yellow Paper*, 2014.

[14] "Ropsten testnet pow chain." https://github.com/ethereum/ropsten, 2020. [Online].